

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 10-005000

(43)Date of publication of application : 13.01.1998

(51)Int.Cl.

C12Q 1/68  
// C12N 15/09

(21)Application number : 08-167770

(71)Applicant : HITACHI LTD

(22)Date of filing : 27.06.1996

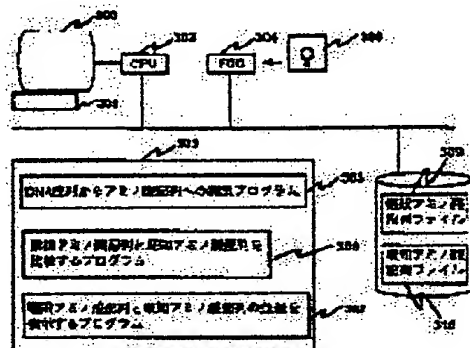
(72)Inventor : KASAHARA NAOKO  
NAGAI KEIICHI  
HIRAOKA SUSUMU

## (54) METHOD FOR COMPARING DNA AMINO ACID SEQUENCE

## (57)Abstract:

**PROBLEM TO BE SOLVED:** To enable the picking of analogous amino acids even in the presence of the insertion or deletion of base units in a DNA sequence by dividing a DNA sequence into base groups of a prescribed length, translating the divided base group into amino acids by shifting the base in a prescribed direction and comparing the translated amino acids with known amino acids.

**SOLUTION:** A program 305 for translating a DNA sequence to an amino acid sequence, a program 306 for comparing the translated amino acid sequence with a known amino acid sequence by Smith-Waterman method and a program 307 for displaying the translated amino acid sequence together with the known amino acid sequence from the compared result are stored in a main memory 303 of a DNA amino acid sequence comparison apparatus. A DNA sequence is divided into base groups of a prescribed length, the divided base group is translated into amino acids while shifting the base in a prescribed direction and the translated amino acid sequence is displayed together with the known amino acid sequence.



## LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision]

of rejection]

[Date of requesting appeal against examiner's  
decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2000 Japanese Patent Office

(19)日本国特許庁(J P)

(12) 公開特許公報(A)

(11)特許出願公開番号

特開平10-5000

(43)公開日 平成10年(1998)1月13日

(51)Int.Cl. <sup>6</sup>	識別記号	庁内整理番号	F I	技術表示箇所
C 1 2 Q 1/68	Z N A	7823-4B	C 1 2 Q 1/68	Z N A Z
// C 1 2 N 15/09		9282-4B	C 1 2 N 15/00	A

審査請求 未請求 請求項の数7 O L (全 11 頁)

(21)出願番号 特願平8-167770

(22)出願日 平成8年(1996)6月27日

(71)出願人 000005108

株式会社日立製作所

東京都千代田区神田駿河台四丁目6番地

(72)発明者 笠原 直子

東京都国分寺市東恋ヶ窪一丁目280番地

株式会社日立製作所中央研究所内

(72)発明者 永井 啓一

東京都国分寺市東恋ヶ窪一丁目280番地

株式会社日立製作所中央研究所内

(72)発明者 平岡 進

東京都国分寺市東恋ヶ窪一丁目280番地

株式会社日立製作所中央研究所内

(74)代理人 弁理士 平木 祐輔

(54)【発明の名称】 DNAアミノ酸配列比較方法

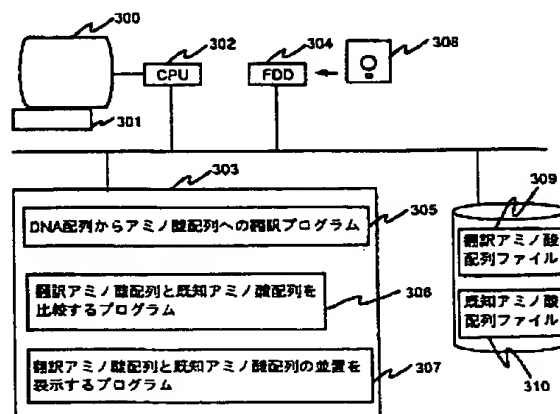
(57)【要約】

【課題】 DNA配列に挿入あるいは欠失が存在しても類似している既知アミノ酸配列を高感度に検索できるDNA配列とアミノ酸配列の直接比較方法を提供する。

【解決手段】 DNA配列をアミノ酸への翻訳規則にのっとってアミノ酸配列に翻訳する手段305、翻訳アミノ酸配列と既知アミノ酸配列をDNA配列に存在する挿入あるいは欠失を考慮して比較を行う手段306、比較結果に基づいて、翻訳アミノ酸配列と既知アミノ酸配列の並置結果をDNA配列と共に表示する手段307を順次実行し、検索結果を出力する。

【効果】 新たに決定されたDNA配列に対して配列中に存在する挿入あるいは欠失を考慮して、類似しているアミノ酸配列を高感度に検索を行うことができ、類似している部分を並置結果として表示することで、DNA配列の持つ機能を類推することが容易となる。

本発明の配列比較方法を適用する配列比較装置の構成を示す図



## 【特許請求の範囲】

【請求項1】 塩基の挿入あるいは欠失を含むまたは含まないDNA配列とアミノ酸配列を直接比較する比較方法において、前記DNA配列を所定の長さの塩基群に分割し、分割された塩基群について所定の方向に塩基をずらしてアミノ酸に翻訳し、そのデーターを前記アミノ酸配列と比較して並置を行いその結果を表示することを特徴とするDNAアミノ酸配列比較方法。

【請求項2】 塩基の挿入あるいは欠失を含むまたは含まないDNA配列とアミノ酸配列を直接比較する比較方法において、前記DNA配列を所定の長さの塩基群に分割し、分割された塩基群について5'又は3'末端から1又は2塩基づつずらしてアミノ酸に翻訳し、そのデーターを塩基の挿入あるいは欠失を考慮したあらゆる組み合わせを想定し、最適経路を選ぶ方法により前記アミノ酸配列と比較して並置を行いその結果を表示することを特徴とする請求項1記載のDNAアミノ酸配列比較方法。

【請求項3】 DNA配列を5'又は3'末端から所定の方向に塩基を順次シフトして、前記塩基群に分割し、分割されたDNA配列から翻訳されたアミノ酸配列と比較の対象となるアミノ酸配列の間で、それぞれのアミノ酸について類似度を積算し、その類似度の積算結果が最大となるように、前記DNA配列を所定の方向に順次シフトし、翻訳アミノ酸配列を選択することを特徴とする請求項1又は2に記載のDNAアミノ酸配列比較方法。

【請求項4】 DNA配列入力手段、前記アミノ酸配列入力手段、DNA配列からアミノ酸配列への翻訳手段、該翻訳アミノ酸配列と比較の対象となる前記アミノ酸配列の配列比較手段、該配列比較手段中で類似度を積算する際に参照するスコアテーブル、該翻訳アミノ酸配列と前記アミノ酸配列を並置して前記DNA配列とともに表示する手段を有することを特徴とする請求項1乃至3のいずれかの項に記載のDNAアミノ酸配列比較方法。

【請求項5】 請求項3に記載のDNAアミノ酸配列比較方法において、DNA配列からアミノ酸配列への翻訳する方法が、DNA配列の5'又は3'末端から3文字単位で1文字ずつずらしながら順次翻訳規則にしたがってアミノ酸配列に翻訳する方法であり、DNA配列から翻訳されたアミノ酸配列と比較の対象となるアミノ酸配列の配列比較プログラムの類似度を積算する方法が、動的計画法を用いるものであり該動的計画法演算でマトリクスの一方の軸を翻訳アミノ酸配列に他方の軸を比較の対象となる前記アミノ酸配列に対応させた時に翻訳アミノ酸配列のi番目の塩基と比較の対象となる前記アミノ酸配列のj番目のアミノ酸塩基の対の類似度を積算する際に、前記DNA配列中の挿入あるいは欠失が存在する場合と比較の対象となる前記アミノ酸配列中に挿入あるいは欠失が存在した場合を考慮して、(1) i-3番目とj-1番目の類似度から積算する場合、(2) i番目

とj-1番目の類似度から積算する場合、(3) i-3番目とj番目の類似度から積算する場合、(4) i-4番目とj-1番目の類似度から積算する場合、(5) i-7番目とj-2番目の類似度から積算する場合、

(6) i-2番目とj-5番目の類似度から積算する場合、(7) i-1番目とj-5番目の類似度から積算する場合の7種類の経路のうちの少なくとも1つの経路を用いて類似度を積算し、該動的計画法を基に翻訳アミノ酸配列と比較の対象となる前記アミノ酸配列間の類似度の積算値と並置を求めることを特徴とするDNAアミノ酸配列比較方法。

【請求項6】 比較の対象となる前記アミノ酸配列が既知アミノ酸配列であることを特徴とする請求項1乃至5のいずれかの項に記載のDNAアミノ酸配列比較方法。

【請求項7】 前記DNA配列が既知DNA配列であることを特徴とする請求項1乃至5のいずれかの項に記載のDNAアミノ酸配列比較方法。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】本発明はDNAアミノ酸配列比較方法に関し、特にDNA配列とアミノ酸配列を比較しDNA配列から直接類似アミノ酸配列を検索することに適したDNAアミノ酸配列比較方法である。

## 【0002】

【従来の技術】近年、様々な生物のDNA配列を決定し、その機能を解析する動きが盛んになっている。DNA配列はA、C、G、Tの4種類の塩基の並びであり、この塩基の並びの一部がそれぞれ生体中で機能する蛋白質をコードしている。重要な機能を持つ蛋白質は製薬などに利用されることが出来るため、DNA配列から直接コードする蛋白質の機能を推定する技術が望まれている。実際に生体中で機能するのは蛋白質配列であるが、DNA配列の決定を行なうほうが直接蛋白質配列を決定するよりも技術的に容易であるため、一般的にはDNA配列を決定する。従って、新たに決定されたDNA配列の機能を推定するには、機能が分かっている蛋白質と比較し、類似しているか否かを判定することになる。

【0003】DNAは3文字の塩基単位(コドン)ごとに20種類のアミノ酸の一つに翻訳される。DNA配列の一部がアミノ酸への翻訳領域であり、その翻訳開始点や終了点の規則は正確には解明されていない。従って、DNA配列に塩基レベルの挿入あるいは欠失といった誤りが存在した場合には、DNA配列中のアミノ酸への翻訳領域がずれてしまうこともある。

【0004】また、DNA配列はコドン単位でアミノ酸に翻訳されるために、上記の様な誤りがおきた場合には全く異なるアミノ酸として翻訳されてしまうことも考えられる。従来の方法ではこのようなDNA配列に塩基の挿入あるいは欠失が生じるような誤りに対応した検索を行っていなかった。従来のDNA配列とアミノ酸配列

を直接比較する方法として、BLASTX (コロナ社: バイオテクノロジー教科書シリーズ11「バイオテクノロジーのためのコンピュータ入門」中村春木・中井謙太共著141P-143P、1996)がある。この方法では図1で表わされる方法で、DNA配列からアミノ酸配列への翻訳を行なう。即ち、まずDNA配列の端からコドン単位でアミノ酸に翻訳するフレーム(1)、コドンの開始位置をフレーム(1)より1文字ずらしてアミノ酸に翻訳するフレーム(2)、コドンの開始位置をフレーム(1)の開始位置よりも2文字ずらしてアミノ酸に翻訳するフレーム(3)、また該DNA配列の相補鎖の反対側の端からコドン単位でアミノ酸への翻訳を開始するフレーム(4)、フレーム(4)の翻訳開始位置から1文字ずらしてアミノ酸への翻訳を開始するフレーム(5)、フレーム(4)の翻訳開始位置から2文字ずらしてアミノ酸への翻訳を開始するフレーム(6)という6つのフレームについてDNA配列をアミノ酸配列に翻訳する。6種類に翻訳されたアミノ酸配列に対して、それぞれ既知アミノ酸配列と比較を行なう。この方法では、DNA配列に塩基単位の挿入あるいは欠失が生じて翻訳フレームが途中でずれてしまった場合に対応していない。例えば、フレーム(1)で翻訳されたアミノ酸配列に非常に類似したアミノ酸配列が存在したとしても、DNA配列中に塩基単位で挿入あるいは欠失が存在した場合には、その場所より後はフレームが(2)あるいは(3)に変更される。しかし、従来方法ではそのようなフレームの変更に対応しきれていない。BLASTXでは、配列の比較検索を6種類の翻訳フレームを利用し、確率計算を行うことで類似配列を類推する方法を用いている。しかし、この方法では検索もれが生じる可能性がある。また、6種類のフレームのそれぞれについて比較を行なっているために、アミノ酸配列への翻訳が次にどのフレームに変更されているのかが分かりにくいという問題点も生じている。

【0005】また第2に従来方法として、Smith-Waterman法 (Identification of Common Molecular Subsequences, J. Mol. Biol., (1981), 147, 195-197, T. F. Smith and M. S. Waterman)がある。この方法は、図2に示したように、比較する2つの配列の文字を1文字ずつ比較して、それぞれに対応したスコアを与え積算し、最終的にスコアが最大となるような検索経路を計算する方法である。この方法は従来ある検索方法の中でもっとも正確な検索方法であるが、配列中の塩基、1文字1文字について比較を行なうために時間がかかる。1組の配列同士のみならず、この場合だと、上記6種類に翻訳されたアミノ酸配列に対してそれぞれ検索を行わなくてはならないために、さらに検索時間がかかる。また、この方法はアミノ酸単位、あるいはDNA配列中のコドン単位

での挿入あるいは欠失には対応できるものの、DNA配列中の塩基単位での挿入あるいは欠失には対応できない。また、この方法でもフレーム間の変更が分かりにくいという問題点も生じる。

#### 【0006】

【発明が解決しようとしている課題】従来のDNA配列、アミノ酸配列の比較検索方法では、DNA配列の方に塩基単位で挿入あるいは欠失が存在した場合には類似アミノ酸配列を拾ってこないという検索もれをおこす可能性がある。Smith-Waterman法では、DNA配列を考える全てのアミノ酸翻訳フレーム6種類について翻訳し、それぞれに翻訳されたアミノ酸配列を用いて配列比較を行なうために、非常に計算時間がかかる上に異なるフレーム間にまたがった場合にうまく類似部分を特定することが困難である。計算を高速化するために開発された従来方法のひとつであるBLASTXでは、確率計算を用いて高速化の実現を行なっているためにさらに検索もれを起こす可能性がある。本発明は、そのようなDNA配列中に存在する塩基単位の挿入あるいは欠失を考慮して、なおかつ検索もれの無いDNA配列とアミノ酸配列の比較を行なうことが可能とするものである。

#### 【0007】

【課題を解決するための手段】本発明の特徴は、以下の処理ステップを含むDNA配列とアミノ酸配列比較方法にある。

【1】DNA配列を塩基単位の挿入あるいは欠失を考慮してアミノ酸配列に翻訳するステップ。

【2】前記DNA配列からの翻訳アミノ酸配列と既知アミノ酸配列を挿入あるいは欠失を考慮しつつ配列比較を行なうステップ。

【3】比較結果に基づいて、翻訳アミノ酸配列と既知アミノ酸配列の並置結果をDNA配列とともに表示するステップ。

【0008】即ち、本発明は塩基の挿入あるいは欠失を含むまたは含まないDNA配列とアミノ酸配列を直接比較する比較方法において、前記DNA配列を所定の長さの塩基群に分割し、分割された塩基群について所定の方向に塩基をずらしてアミノ酸に翻訳し、そのデータを前記アミノ酸配列と比較して並置を行いその結果を表示することを特徴とするDNAアミノ酸配列比較方法である。

【0009】更に、本発明は塩基の挿入あるいは欠失を含むまたは含まないDNA配列とアミノ酸配列を直接比較する比較方法において、前記DNA配列を所定の長さの塩基群に分割し、分割された塩基群について5'又は3'末端から1又は2塩基づつずらしてアミノ酸に翻訳し、そのデータを塩基の挿入あるいは欠失を考慮したあらゆる組み合わせを想定し、最適経路を選ぶ方法により前記アミノ酸配列と比較して並置を行いその結果を表

示することを特徴とするDNAアミノ酸配列比較方法である。

【0010】更に、本発明はDNA配列を5'又は3'末端から所定方向に塩基を順次シフトして、前記塩基群に分割し、前記DNA配列から翻訳された該翻訳アミノ酸配列と比較の対象となる前記アミノ酸配列の間で、それぞれのアミノ酸について類似度を積算し、その類似度の積算結果が最大となるように、前記DNA配列を所定方向に順次シフトし、該翻訳アミノ酸配列を選択する事を特徴とする前記DNAアミノ酸配列比較方法である。

【0011】更に、本発明はDNA配列入力手段、前記アミノ酸配列入力手段、DNA配列からアミノ酸配列への翻訳手段、該翻訳アミノ酸配列と比較の対象となる前記アミノ酸配列の配列比較手段、該配列比較手段中で類似度を積算する際に参照するスコアテーブル、該翻訳アミノ酸配列と前記アミノ酸配列を並置して前記DNA配列とともに表示する手段を有することを特徴とする前記DNAアミノ酸配列比較方法である。

【0012】更に、本発明は前記DNAアミノ酸配列比較方法において、DNA配列からアミノ酸配列への翻訳する方法が、DNA配列の5'又は3'末端から3文字単位で1文字ずつずらしながら順次翻訳規則にしたがってアミノ酸配列に翻訳する方法であり、DNA配列から翻訳されたアミノ酸配列と比較の対象となるアミノ酸配列の配列比較プログラムの類似度を積算する方法が、動的計画法を用いるものであり該動的計画法演算でマトリクスの一方の軸を翻訳アミノ酸配列に他方の軸を比較の対象となる前記アミノ酸配列に対応させた時に翻訳アミノ酸配列のi番目の塩基と比較の対象となる前記アミノ酸配列のj番目のアミノ酸塩基の対の類似度を積算する際に、前記DNA配列中の挿入あるいは欠失が存在する場合と比較の対象となる前記アミノ酸配列中に挿入あるいは欠失が存在した場合を考慮して、(1) i-3番目とj-1番目の類似度から積算する場合、(2) i番目とj-1番目の類似度から積算する場合、(3) i-3番目とj番目の類似度から積算する場合、(4) i-4番目とj-1番目の類似度から積算する場合、(5) i-7番目とj-2番目の類似度から積算する場合、

(6) i-2番目とj-5番目の類似度から積算する場合、(7) i-1番目とj-5番目の類似度から積算する場合の7種類の経路のうちの少なくとも1つの経路を用いて類似度を積算し、該動的計画法を基に翻訳アミノ酸配列と比較の対象となる前記アミノ酸配列間の類似度の積算値と並置を求めることを特徴とする前記DNAアミノ酸配列比較方法である。

【0013】上記比較の対象となるアミノ酸配列としては、例えばアミノ酸配列データベースから選択された既知アミノ酸配列を用いることができる。上記DNA配列としては、例えばDNAデータベースから選択された

既知DNA配列を用いることができる。

【0014】

【発明の実施の形態】

【0015】

【実施例】本発明の第1の実施例について図3を用いて説明する。本実施例はディスプレイ300、キーボード301、中央演算装置CPU302、主メモリ303、フロッピーディスクドライブ304から構成される。主メモリ303には、DNA配列からアミノ酸配列への翻訳プログラム305、翻訳アミノ酸配列と既知アミノ酸配列を比較するプログラム306、比較した結果から翻訳アミノ酸配列と既知アミノ酸配列の並置を表示するプログラム307が格納されている。これらのプログラムはCPU302で実行される。

【0016】DNA配列登録の際には、キーボード301から入力されたコマンドにより、CPU302がフロッピーディスクドライブ304に挿入されるフロッピーディスク308からDNA配列を読み取り、DNA配列からアミノ酸配列への翻訳プログラム305を実行して作成された配列を翻訳アミノ酸配列ファイル309として格納する。既知アミノ酸配列は、DNA配列と同様にフロッピーディスク308から読み込むか、あるいは既に登録されていたアミノ酸配列データベースから読み込み、既知アミノ酸配列ファイル310として格納する。

【0017】配列比較の際には、CPU302が翻訳アミノ酸配列ファイル309と既知アミノ酸配列ファイル310から配列を読み込んで、翻訳アミノ酸配列と既知アミノ酸配列を比較するプログラム306を実行する。更に、実行結果を用いて、翻訳アミノ酸配列と既知アミノ酸配列の並置を表示するプログラム307を実行し、配列比較として出力する。以上が本発明のDNA配列とアミノ酸配列の直接比較方法を実現するシステムである。

【0018】以下に、DNA配列に塩基単位の挿入あるいは欠失が存在することを考慮して、アミノ酸配列に翻訳するプログラム305について説明する。図4に示したように、DNAは3文字単位のコドン毎に1種類のアミノ酸にコードされる。コドンは4種類のDNA塩基が3つ組み合せて決定されるものなので、64種類のコドンがあり得る。ところがアミノ酸は20種類でしかないので、複数のコドンが一つのアミノ酸をコードしていることとなる。このコドンのアミノ酸へのコード規則、すなわち、アミノ酸への翻訳規則を示したのが図4である。

【0019】つぎに、この翻訳規則を用いてDNA配列を仮想的にアミノ酸配列に翻訳する。これは、DNA配列とアミノ酸配列を直接比較する際に、まず、DNA配列を計算上仮想的にアミノ酸配列に翻訳し、そのようにして翻訳されたアミノ酸配列と実際のアミノ酸配列とを比較する方法を用いているからである。従って、図5に

示した方法でDNA配列をアミノ酸配列に翻訳する。つまりDNA配列の端からコドンを取り出し該当するアミノ酸に翻訳、次に1文字ずらしてコドンを取り出して同様に該当するアミノ酸に翻訳する。この動作をくり出したコドンの最後の文字が、DNA配列の最後の文字になるまでくり返し、最終的にDNA配列をアミノ酸配列に翻訳する。図5の例では、DNA配列がATGCA・・・・CGATなので、まず端から最初のコドンATGを取り取り対応するアミノ酸Mに翻訳する。翻訳アミノ酸配列の1文字目はMとなる。次にDNA配列から翻訳アミノ酸配列の2文字目に当たるコドンTGCを、前のコドン開始位置から1文字ずらして切り出しアミノ酸Cに翻訳する。更にDNA配列の1文字ずらした位置からコドンGCAを切り出しアミノ酸Aに翻訳する。この動作をくり返し、DNA配列からアミノ酸配列を翻訳する。図5の例の場合には、翻訳されたアミノ酸配列はMCA・・・・RDとなる。このDNA配列からアミノ酸配列への仮定の翻訳は、通常のDNA配列からアミノ酸配列に比較して約3倍量のアミノ酸配列が翻訳されることとなる。この方法を用いることにより、DNA配列は1本のアミノ酸配列、相補鎖を考慮してもせいぜい2本のアミノ酸配列に翻訳される。このようにして翻訳されたアミノ酸配列と既知アミノ酸配列を、Smith-Waterman法をもとにDNA塩基単位での挿入あるいは欠失を許容する配列比較方法にて比較する事により、その配列間の類似度を見る事が出来る。

【0020】以下に、翻訳アミノ酸配列と既知アミノ酸配列間の配列比較のプログラム306について詳しく説明する。本発明は、Smith-Waterman方法に基づいた配列比較方法を用いている。図6に示されたアミノ酸同士の対に対するスコア表を用いて、翻訳アミノ酸配列と比較対象となっている既知アミノ酸配列の間のスコアを算出し、算出されたスコアに応じてその配列の類似度を類推する方法である。このスコアマトリクスは、アミノ酸の各組の性質の類似度を考慮して、それぞれの組に対してスコアを設定するものである。このマトリクスの値は、どの位の類似度の配列を検索することによって、検索者自身が選択することが可能である。図6は、いくつか実際に使用されているマトリクスの中で、blosun62という名前のマトリクスである。

【0021】次に、本発明における配列比較方法の中心部分を占めるスコアの算出方法について、図7を用いて説明する。まず、図7に表わされるように翻訳アミノ酸配列と既知アミノ酸配列をマトリクス状に配置する。配置された配列の各要素に対応するアミノ酸同士のスコアは、図6に示されたスコアマトリクスを参照する。配列の要素であるアミノ酸を読み込み、挿入あるいは欠失に対するペナルティを与えながら、スコアを積算していく事でマトリクス内の升を埋めて行く。しかし、この積算時に参照するスコアは図7に示された7つの場合から積

算して行く。即ち、図7の(0)の位置のスコアを計算するには、(1)から(7)の位置のスコアに、(0)の位置のアミノ酸対のスコアを図6のマトリクスから参照し加算または適宜に挿入あるいは欠失に対するペナルティスコアを加算して、結果が最大値をとるようなスコアを選択する。この時、(1)から(7)までのどの部分のスコアに加算されたのかを記録しておく。図8に示してあるように、(1)に加算されたスコアが(0)におけるスコアの最大値である時、この場合は翻訳アミノ酸配列と既知アミノ酸配列のアミノ酸が一致していても、不一致であっても対応させる場合である。即ち、翻訳する前のDNA配列にも既知アミノ酸配列にも挿入あるいは欠失をいれない場合である。(2)に加算されたスコアが最大値となる時は、翻訳アミノ酸配列にアミノ酸1文字の欠失がある場合、即ち、翻訳する前のDNA配列の該当する部分にコドン単位の欠失が存在する場合である。(3)に加算されたスコアが最大値となる時は、既知アミノ酸配列に1文字の欠失が存在する場合である。(4)に加算されたスコアが最大値となる時は、(0)の部分の翻訳アミノ酸に対応するDNA配列中のコドンの直前に1塩基の挿入がある場合である。(5)に加算されたスコアが最大値となる時は、(0)の部分の翻訳アミノ酸の直前の翻訳アミノ酸に対応するDNA配列中のコドンの中に塩基が挿入している場合である。(6)に加算されたスコアが最大値となる時は、(0)の部分の翻訳アミノ酸の直前の翻訳アミノ酸に対応するDNA配列中のコドン中の塩基が欠失している場合である。(7)に加算されたスコアが最大値となる時は、

(0)の部分の翻訳アミノ酸の直前の翻訳アミノ酸に対応するDNA配列中のコドン単位の挿入とそのコドン中の塩基が欠失している場合である。以上の7つの場合を考慮してスコアの計算を行なう。実際のスコア算出方法に関しては、次に示す。翻訳アミノ酸配列のi番目と既知アミノ酸配列j番目のスコアs(i,j)を求める式s(i,j)=max[score0,score1,score2,score3,score4,score5,score6,score6]

$$\text{score } 0 = \text{score}(i, j) + s(i-1, j-3)$$

$$\text{score } 1 = s(i-1, j) - 4 \text{ or } s(i-1, j) - 12$$

$$\text{score } 2 = s(i, j-3) - 4 \text{ or } s(i, j-3) - 12$$

$$\text{score } 3 = s(i-1, j-4) + \text{score}(i, j) - 12$$

$$\text{score } 4 = s(i-2, j-7) + \text{score}(i, j) - 12$$

$$\text{score } 5 = s(i-2, j-5) + \text{score}(i, j) - 12$$

$$\text{score } 6 = s(i-1, j-7) + \text{score}(i, j) - 24$$

$$\text{score } 7 = s(i-1, j-5) + \text{score}(i, j) - 24$$

score(i,j):i番目の塩基とj番目の塩基対に与えられる類似度の指標

式中の減算されている数字(例 -4, -12, -24)は挿入・欠失またはその延長に対して与えられるペナルティスコアである。

【0022】上記で説明した方法に基づき、翻訳アミノ

酸と既知アミノ酸配列とのスコアを算出する。マトリクスの端でスコアが最大値をとる升を選択し、その最大スコアを配列間の類似スコアとし、配列比較の結果を示す指標とする。このスコアが大きいほどより類似しているDNA配列とアミノ酸配列であると言う事が出来る。この結果に基づき、翻訳アミノ酸配列と既知アミノ酸配列の並置を表示する。

【0023】以下に、翻訳アミノ酸配列と既知アミノ酸配列の並置の表示プログラム307について、図8を用いて説明する。上記で説明したように、図7で示された(1)から(7)のどの部分のスコアを参照してスコアが算出されたかによって、挿入あるいは欠失の存在する位置が変化する。翻訳アミノ酸配列と既知アミノ酸配列がマトリクス状に配置され、各アミノ酸に対応するスコアが算出された後、配列の一番端を示すマトリクスの行および列上でスコアが最大値をとる位置から、その部分のスコアが図7の(1)から(7)のどの場所から計算されたかをスコアが0になるまでたどって行く。それぞれ(1)から(7)のどの場合でも、図8に対応する並置例をつなげていく事で、最終的に翻訳アミノ酸配列と既知アミノ酸配列の並置を求め、表示する。また、アミノ酸は1文字表記の他に3文字表記も一般的になされているので、翻訳アミノ酸配列と既知アミノ酸配列とを1文字表記で表現するのではなく3文字表記で表示する事で、翻訳アミノ酸配列とDNA配列をならべて表示する事が出来る。この時、DNA配列には図7および図8の規則に従い、挿入あるいは欠失を示す記号を該当箇所に代入する事で、既知アミノ酸配列とDNA配列との比較をより分かりやすく表示する事が可能である。

【0024】以下、本発明による比較手順を図9に従って、実際のDNA配列とアミノ酸配列を用いて説明する。図9中の901に示したように、DNA配列をagcttgccaaactとする。図5中で説明した手順に従い、すなわち、DNA配列の片方の端から1文字づつずらしながらコドン単位でアミノ酸に翻訳する。コドンがアミノ酸に翻訳される規則は図4に示されている。この規則を用いて上記agcttgccaaactというDNA配列は、図9の902に示したようにまず1番端のコドンagcがアミノ酸Ser(1文字表記ではS)に翻訳され、次にgctがアミノ酸Ala(A)に翻訳される。このように1文字づつずらしながらコドン単位でアミノ酸に翻訳していくという操作を繰り返し、上記DNA配列は図9中903のアミノ酸配列SALLCAPQNTに翻訳される。この903のアミノ酸配列を比較対象となるアミノ酸配列と区別するために翻訳アミノ酸配列と呼ぶことにする。次にこの様にして作成された翻訳アミノ酸配列と、データベース中などの既知のアミノ酸配列との比較方法を説明する。翻訳アミノ酸配列903を既知アミノ酸配列904と比較する場合を例にとって説明する。図9中904の既知アミノ酸配列SARA

PQRDTと903の翻訳アミノ酸配列SALLCAPQNTを比較する場合には以下の手順に従う。まず、905に示すように翻訳アミノ酸配列を垂直方向、既知アミノ酸配列を水平方向に配置したマトリクスを考える。基本的な配列比較の方法は、この様にして作成されたマトリクス内の全てのマスにおける類似度の基準となるスコアを算出し、その最大スコアによって、類似しているかしていないかの判別を行う。スコアの算出方法を説明する。図6に示されたようにそれぞれのアミノ酸対には、類似度の指標であるスコアが与えられる。このスコア体系は、求める進化上の距離に応じて選択することができるが、ここでは図6に示したスコアマトリクスB1osum62を用いる。マトリクス上の各マスにおけるスコアは図7に示したように基本的には既に計算して求められた7つのマスのスコアから算出し、その最大値を選択することによって、該当するマスにおけるスコアを計算する。まず、一番上の行のスコアを計算する。この行は、翻訳アミノ酸配列の一番最初のアミノ酸であるSと、水平方向に配置された既知のアミノ酸配列904のSARAPQRDTとの間のスコアを算出する。スコアは図に示された7つのマスのスコアから算出されるが、この行のように7つのマスのスコアがまだ計算されていない場合には、スコアの初期値は0として計算を行う。まず、2つの比較する配列903と904の一番最初のアミノ酸SとSの対に与えられる値は、図6のスコアマトリクスを参照して4であることが分かる。従って、この4という値を図7に示された7つのマスのスコアに加算して、それぞれに得られた値のうち最大値をスコアとする。そのため、図9中の905のマトリクスの1行目の最初マス906のスコアは4となる。次に、翻訳アミノ酸配列903中の1番目のアミノ酸Sと既知のアミノ酸配列904中の2番目のアミノ酸Aとの比較スコア、すなわち、905のマトリクス内の1行2列目のマス907に該当するスコアを算出する。このマス907のスコアは図に示された7つのマスのスコアに、アミノ酸SとAの対に与えられる値1を加算し、その最大値を選択することで求める。ここで、図7に示された7つのマスのうち、(2)に対応するマス906以外はスコアが求められていない。従って、このマスのスコアは

(2)に対応するマス906、すなわち翻訳アミノ酸配列903の1番目のアミノ酸Sと既知のアミノ酸配列904の1番目のアミノ酸Sとの比較で算出されたスコア4に、アミノ酸SとAの対に対するスコア1を加算して5という値を得、その値が他の場合から算出される1という値よりも大きいので、このマス907におけるスコアは5となる。マトリクス905の1行目は図7における(2)に対応するマスの値のみを参照して、スコアを算出することになるが、以下、行を重ねるに従って図7に示された7つのマスのスコアを参照して、スコア計算を行うこととなる。この操作を繰り返して、それぞれの



マスに対応するスコアを算出する。マトリクス905のマススコアを全て計算しものが908である。マトリクス908の各マス内の円で囲まれた数字がそのマスにおけるスコアであり、左上にある数字は、そのマスにおけるスコアが図7の(1)から(7)までのどのマスのスコアから算出されたかを示す数字である。そして908のマトリクス上の1番端の行及び列上において最大値を探しその値を翻訳アミノ酸配列903と既知のアミノ酸配列904の配列比較におけるスコアとなる。このスコアの大小によって、既知のアミノ酸配列904が翻訳アミノ酸配列903に類似しているか否かを判断する基準とする。この例の場合では、マトリクス908の垂直成分に当たる翻訳アミノ酸配列903の最後のアミノ酸Tの行と、水平成分に当たる既知のアミノ酸配列904の最後のアミノ酸Tの列における最大スコアをこの配列比較に対するスコアとする。次に、このスコアの計算結果から翻訳アミノ酸配列903と既知のアミノ酸配列904の間の並置を求める手順を説明する。並置は、比較が行われた配列間で、配列のどの部分がどのように類似しているかを表示する方法である。並置は最大スコアに対応するマスから、図7の(1)から(7)のどのマスからそのスコアが算出されたかをたどり、(1)から

(7)の経路に従って図8のような並置例を繋げていく事によって求められる。この例の場合には、まずマトリクス908の最大スコアをとるマス909からたどっていくこととなる。マス909のスコアは、図7における(2)にあたるマス910から計算されているので、909から910へと並置経路をたどる。マス910のスコアも同様に図7の(2)にあたるマスから計算されているので、マス911に並置経路をたどる。マス911のスコアは、図7の(1)にあたるマス912から計算されているので、並置経路は911から912に飛ぶことになる。このような手順をくり返し、配列を比較した結果の並置経路を求めることが出来る。求められた並置経路に対して、それぞれのマスのスコアが図の(1)から(7)のいずれかのマスのスコアから計算されたかに従って、図8の並置例に従い、並置結果を表示することが出来る。従って、翻訳アミノ酸配列903、すなわちDNA配列901と、既知のアミノ酸配列904の比較結果としての並置の表示は、913に示したようになる。図9で説明に用いた例は、配列が非常に短いため、挿入・欠失に対するペナルティを考慮して計算すると、スコア自身が非常に小さな値になってしまい、検索が出来なくなる。そのため、ここでは配列比較方法の原理を説明するために、挿入・欠失に対するペナルティは考慮しなかったが、実際の検索の時には、[数1]に表わされるようなペナルティを導入して、スコア計算を行っている。これはもし、挿入・欠失に対してペナルティを導入しないと、無制限に挿入・欠失をいれてしまうことで、実際には類似していない配列を検索で拾ってきて

しまうためである。

【0025】次に上記で説明された本発明の配列比較方法を用いた検索について述べる。本配列比較方法において、実際に配列をもちいた比較を行なう。以下、アミノ酸配列データベースとして、PIR (Release 34) の中でsuperfamily分類の記載のあるデータpir1.seq (配列数10550、アミノ酸塩基数3591370) を利用した。前記既知アミノ酸配列は、アミノ酸配列データベース中に登録されている配列とし、データベース中に含まれているアミノ酸配列のうち実際に翻訳される部分のDNA配列が分かっているDNA配列に、配列塩基長の3%にあたる数の塩基の挿入あるいは欠失を生じさせたものをキーDNA配列として利用した。この値は、実際に解析されたばかりの配列には最悪の場合に3%程の誤りが含まれることを考慮して設定した。従来方法との比較方法としては、元のDNA配列に対応するアミノ酸配列と類似していると分類されている同じsuperfamilyのメンバーをいかに検索で拾ってこれるか否か、または欠失等を生じさせる以前のDNA配列からの翻訳アミノ酸配列に対して、正しい位置に挿入あるいは欠失を考慮してその並置を求める事が出来るかを評価した。図10は、従来方法との比較、即ち、どれだけ同じsuperfamilyのメンバーを認識できるかを示したグラフである。縦軸に実際に検索で拾ってきたsuperfamilyのメンバー数を示しており、横軸には検索で拾ってきたすべての配列数を示している。この結果、配列長の3%の挿入あるいは欠失が存在した場合でも、本発明では従来法よりも効率的に類似したアミノ酸配列を拾ってきている事が分かる。また、図11には、実際に検索を行った並置結果を示している。1行目が既知アミノ酸配列、2行目にはアミノ酸の一致あるいは不一致等をあらわす記号、3行目には翻訳アミノ酸配列、4行目にはDNA配列を配置している。各配列中に現れる記号「-」は挿入あるいは欠失がその位置に存在し、該当する塩基が無い事を表わしている。図11で示されたように、DNA配列中に挿入あるいは欠失が存在している場合でも考慮して検索を行っている事が分かる。図10及び図11に示された結果より、本発明によるDNA配列とアミノ酸配列の比較方法は、類似しているアミノ酸配列を従来方法よりも正確に検索できる事が分かる。

#### 【0026】

【発明の効果】本発明により、新たに決定されたDNA配列に対して、塩基単位の挿入あるいは欠失を考慮してアミノ酸配列と配列比較を行うことができる。その結果、生体中の機能の解明されているアミノ酸配列に類似しているDNA配列を見つけ、その類似部分を並置結果として表示することが出来るために、DNA配列の持つ機能を類推する事が容易になる。

#### 【0027】

## 【配列表】

配列番号: 1  
 配列の長さ: 232  
 配列の型: 核酸  
 鎖の数: 一本鎖  
 トポロジー: 直鎖状  
 配列の種類: mRNA

配列の特徴: HUMROSMCF

トランスメンブラン・プロテインキナーゼの3・末端を  
 コードするヒトmcf3(再配列 ros1)プロト-オンコージ  
 ンmRNA

配列:

【0028】

```

AspPheTrpIleProGluThrSerPheIleLeuThrIleIleValGlyIlePheLeuVal
GATTTTGGATACCGAAACAAGTTTCATCTTACTATTATAGTTGGAATTTCTGGTT
ValThrIleProLeuThrPheVal - - ArgArgLeuLysAsnGlnLysSerAlaLys
GTTACAATCCCACTGACCTTTGTC-----TAGAAGATTAAAGAATCAAAAAGTGCCAAG
GluIleLysValAlaValPro LysThrLeuLysLysGlySerThrAspGlnGluLysIle
GAAATCAAAGTAGCAGTGCCCC-CAAGACTTTGAAGAAGGTTCCACAGACCAGGAGAAGATT
LeuThrLeuValAspLeuValAspLeuCysValAspIle - - GlyCysValTyrLeu
CTCACCTTGGTTGACCTTGAGACCTGTGTAGATATA-----AGGCTGTGTCTACTTG
  
```

【0029】配列番号: 2

配列の長さ: 79  
 配列の型: アミノ酸  
 鎖の数: 一本鎖  
 トポロジー: 直鎖状配列の種類: ペプチド

配列の特徴: TVHURT # タイププロテイン(断  
 片)プロテインチロシンキナーゼmcf3(活性型 ros-1)  
 ヒト断片# EC-No 2.7.1.112

配列:

【0030】

```

AspPheTrpIleProGluThrSerPheIleLeuThrIleIleValGlyIlePheLeuVal
ValThrIleProLeuThrPheValTrpHis ArgArgLeuLysAsnGlnLysSerAlaLys
GluIleLysValAla - - ValLysThrLeuLysLysGlySerThrAspGlnGluLysIle
LeuThrLeuValAspLeuValAspLeuCysValAspIleSerLys GlyCysValTyrLeu
  
```

## 【図面の簡単な説明】

【図1】従来方法におけるDNA配列からアミノ酸配列への翻訳フレームを示す図。

【図2】Smith-Waterman法での配列比較を行なう際のスコアの算出経路を示す図。

【図3】本発明の配列比較方法を適用する配列比較装置の構成を示す図。

【図4】コドンとアミノ酸の対応表を示す図。

【図5】本発明におけるDNA配列からアミノ酸配列への翻訳方法を示す図。

【図6】アミノ酸同士の対に与えられるスコアの一例を示す図。

【図7】本発明におけるスコア算出時の参照位置を示す図。

【図8】本発明における各スコア参照位置からの経路に対して与えられる並置例を示す図。

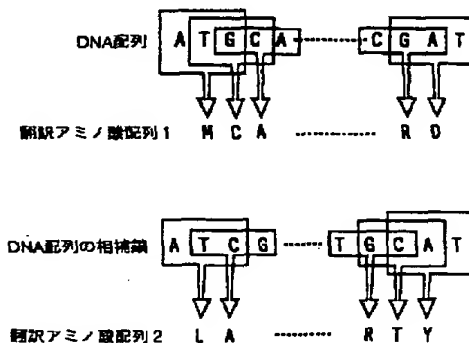
【図9】実例を用いた本発明の配列比較の説明図。

【図10】従来方法との配列比較結果の評価を示す図。

【図11】本発明における配列比較の並置結果を示す図。

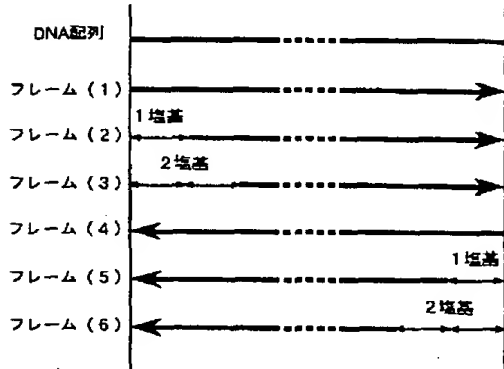
【図5】

本発明におけるDNA配列からアミノ酸配列への翻訳方法を示す図



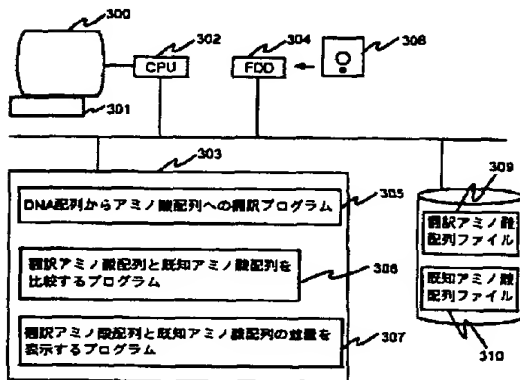
【図1】

従来方法におけるDNA配列からアミノ酸配列への  
翻訳フレームを示す図



【図3】

本発明の配列比較方法を適用する配列比較装置の構成を示す図



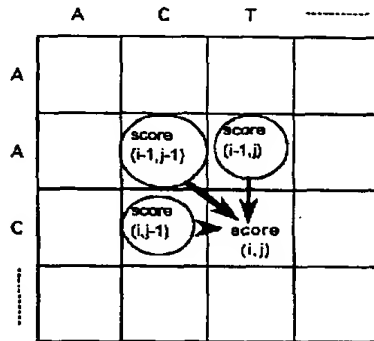
【図6】

アミノ酸同士の対に与えられるスコアの一列を示す図

A	R	M	D	C	Q	E	G	H	I	L	K	F	P	S	T	W	Y	V	B	Z	X	O
Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val	Asx	***	Stop
-4	-1	-2	-2	0	-1	-1	0	-2	-1	-1	-1	-2	-1	0	-3	-2	0	-2	-1	0	-4	
-1	5	0	-2	-3	1	0	-2	0	-3	-2	2	-1	-3	-2	-1	-3	-2	-3	-1	0	-1	
-2	0	6	1	-3	0	0	0	1	-3	-3	0	-2	-3	-2	1	0	-4	-3	3	0	-1	
-2	2	1	6	-3	0	2	-1	-1	-3	-4	-1	-3	-5	-1	0	-1	-4	-3	4	1	-1	
0	-3	-3	-3	9	-3	-4	-3	-3	-1	-3	-1	-2	-3	-1	-1	-2	-1	-3	-3	-2	-4	
-1	1	0	0	-3	5	2	-2	0	-3	-2	1	0	-5	-1	0	-1	-2	-2	0	3	-4	
-1	0	0	2	-4	2	5	-2	0	-3	-3	-1	-2	-5	-1	0	-1	-3	-2	1	4	-1	
-0	-2	0	-1	-3	-2	-2	6	-2	-4	-4	-2	-3	-3	-2	0	-2	-2	-3	-3	-1	-2	
-2	0	1	-1	-3	0	0	-2	8	-3	-3	-1	-2	-1	-2	-1	-2	-2	-3	0	0	-1	
-1	-3	-3	-3	-1	-3	-3	-4	-3	4	2	-3	1	0	-3	-2	-1	-3	-1	-3	-3	-1	
-1	-2	-3	-4	-1	-2	-3	-4	-3	2	4	2	2	0	-3	-2	-1	-2	-1	-4	-3	-1	
-1	2	0	-1	-3	1	1	-2	-1	-3	-2	5	-1	-3	-1	0	-1	-3	-2	0	1	-1	
-1	-1	-2	-3	-1	0	-2	-3	-2	1	2	1	5	0	-2	-1	-1	-1	-1	-3	-1	-1	
-2	-3	-3	-3	-2	-3	-3	-3	-1	0	0	-3	0	6	-4	-2	-2	1	3	-1	-3	-1	
-1	-2	-2	-1	-3	-1	-1	-2	-2	-3	-3	-1	-2	-4	7	-1	-1	-4	-3	-2	-2	-1	
1	-1	1	0	-1	0	0	-1	-2	-2	0	0	-1	-2	1	4	1	3	-2	0	0	4	
0	-1	0	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	5	-2	0	-1	-1	0	-1	-4	
-3	-3	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	
-2	-2	-3	-2	-1	-2	-3	-2	-3	-1	-2	-1	1	1	1	3	-2	11	2	-3	-4	-4	
-0	-3	-3	-3	-1	-2	-2	-3	-3	3	1	-2	1	-1	-2	-2	2	7	-1	-3	-2	-1	
-2	-1	3	4	-3	0	1	-1	0	-3	-4	0	-3	-2	0	-1	-4	-3	-3	4	1	-1	
-1	0	0	1	-3	3	4	-2	0	-3	-3	1	-1	-3	-1	0	-1	-3	-2	1	4	-1	
0	-1	-1	-1	-2	-1	-1	-1	-1	-1	-1	-1	-1	-2	0	0	-2	-1	-1	-1	-1	-4	
-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	-4	

【図2】

Smith-Waterman法での配列比較を行う際の  
スコアの算出経路を示す図



【図4】

コドンとアミノ酸の対応表を示す図

アミノ酸 (三文字表記)	DNA	アミノ酸 (三文字表記)	DNA
A (Ala)	GCN	K (Lys)	AAR
R (Arg)	MGN	M (Met)	ATG
N (Asn)	AAY	F (Phe)	TTY
D (Asp)	GAY	P (Pro)	CCN
C (Cys)	TGY	S (Ser)	WSN
Q (Gln)	CAR	T (Thr)	ACN
E (Glu)	GAR	W (Trp)	TGG
G (Gly)	GGN	Y (Tyr)	TAY
H (His)	CAY	V (Val)	GTN
I (Ile)	ATH	終止	TRR
L (Leu)	YTN	X (xxx)	NNN
B (N or D)	RAY	Z (Q or E)	SAR

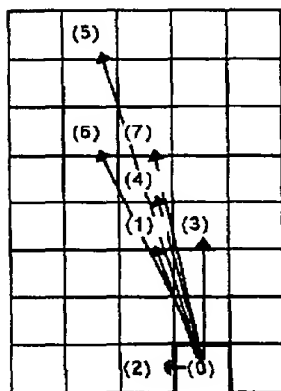
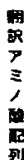
DNA文字種類対応表

R	A/G	M	A/C
Y	C/T	B	C/G/T
S	C/G	V	A/C/G
W	A/T	D	A/G/T
N	A/C/G/T	H	A/C/T

【圖 7】

本発明におけるスコアの参照位置を示す図

既知アミノ酸配列



【图 8】

各スコア参照位置からの経路に対して与えられる並置例を示す図

(1) DNA配列にもアミノ酸配列にも挿入・欠失を入れない

一致の場合：

不一致でスコアが正の場合：

Met	Ala	Asp
Met	Ala	Asp
ATG	GCT	GAC

Met	Ser	Asp
Met	Ala	Asp
ATG	GCT	GAC

不一致でスコアが0:

不一致でスコアが負：

Met	Ser	Asp
Met	Ala	Asp
ATG	GCT	GAC

Met	Hic	Asp
Met	Alm	Asp
ATG	GCT	GAC

(2) DNA配列に3塩基(コドン)単位の欠失 (3)アミノ酸配列に欠失

Leu	Gly	Gln
I		I
Leu	-	Gln
CYT	---	CAS

Gly	-	Gln
Gly	Pro	Gln
GGG	CCC	CAA

#### (4) 比較する翻訳アミノ酸の直前に DNA塩基の挿入

(5)比較する翻訳アミノ酸の直前のコドン中に塩基が挿入

Arg	Ser	Leu
Arg	Ser	Leu
AGA	TCT	TTG

Ala	Asp	G
Ala		G
GCT	GTA	G

(6) 比較する翻訳アミノ酸の  
コドン中の塩基の欠失

(7) 比較する翻訳アミノ酸の直前に  
コドン単位挿入と塩基単位の欠失

Val	Met	Arg
Val		Arg
GTA	A-G	AGA

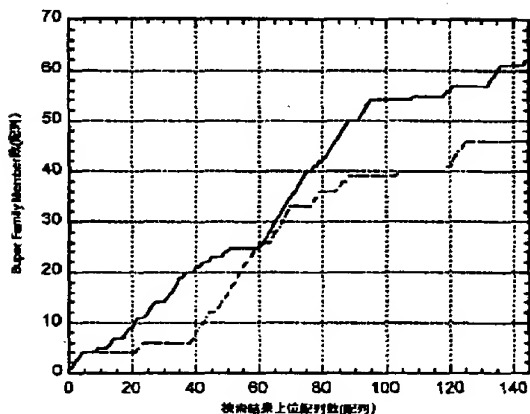
Met -  
Met  
ATB C-C

【図 10】

従来方法との配列比較結果の評価を示す図

—— 本発明による検索  
 ---- BLASTX(従来方法)による検索

**Super Family Number 263**



【图 1-1】

本発明における配列比較の並置結果を示す図

Query sequence:>HUMROSOF (GenBank) Human acf3 (rearranged ros1)  
proto-oncogene mRNA encoding a transmembrane protein kinase, 3' end.

Target sequence:>TYK2 Protein-tyrosine kinase scf3 (activated ras-1) - Human (fragment) #EC-number 2.7.1.112 | 263.0 22.0 2.0 1.0 2.0

Score: 2333 + strand

Target: AspPheTrpIleProGluThrSerPheIleLeuThrIleIleValGlyIlePheLeuVal

Query: AspPheTrpIleProGluThrSerPheIleLeuThrIleIleValGlyIlePheLeuVal  
GATTTTGCATACCAGAACAGTTTCATACTTACTATTATAGTTGGATATTCTGGT

Target: ValThrIleProLeuThrPheValTrpHis ArgArgLeuIleAsnGlnLysSerAlaLys

Query: ValThrIleProLeuThrPheVal - - ArgArgLeuLysAsnGlnLysSerAlaLys  
 GTTAAATCCCACTGACCTTTTCT-----TAAAGATTAAAGATCAAAAAGTCCAG

Target: GluIleLysValAla - - ValLysThrLeuLysLysGlySerThrAspGlnGluLysIle

Query: GluIleLysValAlaValPro LysThrLeutylysGlySerThrAspGlnGluLysIle  
 PAAATCAAGTAGCTHCTCCCTTCAAGACTTTAAGAGAAGGTTTCACAGACCCAGGAGAAAT

Target: **LeuThrLeuValAspLeuValAspLeuCysValAspIleSerLys GlyCysValTyrLeu**

Query: leuThrLeuValAspLeuValAspLeuCysValAspIle - - GlyCysValTyrLeu  
CICACCTTGGTTGACCTTGTAGACCTTGTGTAGATATA-----AGCTGTGTCTACTT

【図9】

